Bioshare Documentation

Release 1.0.1

Adam Schaal

Jul 26, 2017

Contents:

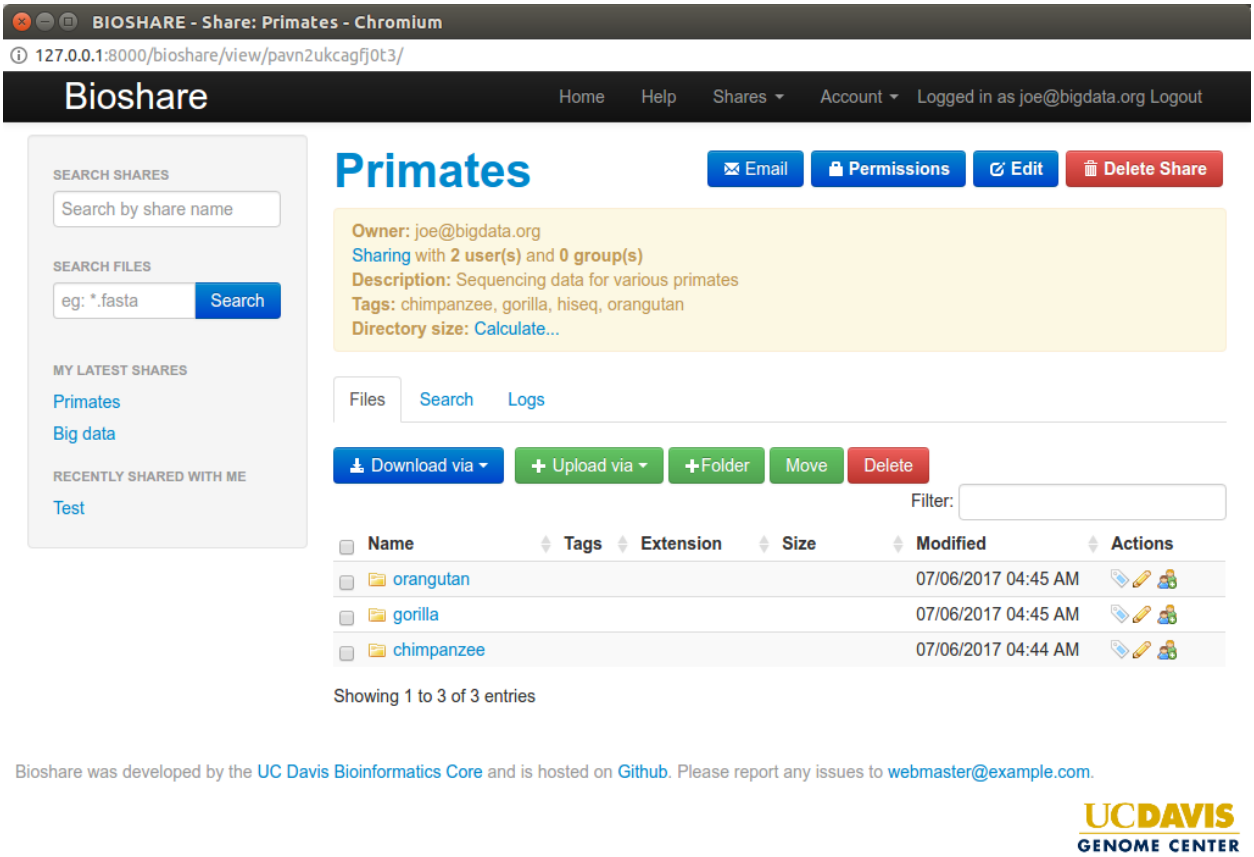
1 About Bioshare **1**

2 User Guide **3**
- 2.1 Home . . . 3
- 2.2 Share . . . 4
- 2.3 Creating a share . . . 6
- 2.4 Share permissions . . . 7
- 2.5 Downloading/Uploading . . . 8
- 2.6 SFTP . . . 9
- 2.7 Rsync (Linux/Mac) . . . 10

3 Indices and tables **13**

CHAPTER 1

About Bioshare



Bioshare is a file sharing web application designed for sharing the "Big Data" common to bioinformatics. While it was developed to support the effective sharing of larger datasets, it can and has been used to share all kinds of data, large and small.

The central goals of Bioshare are:

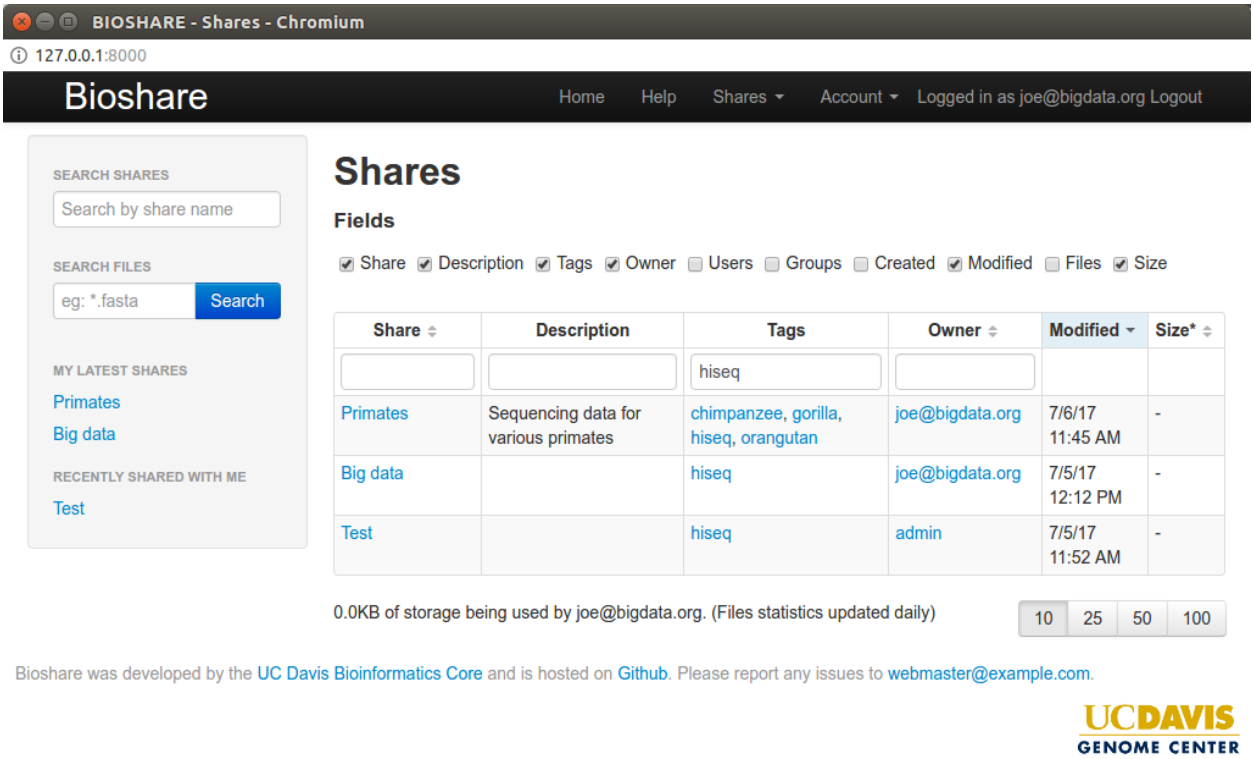
- To easily share data with any collaborator. The only thing that is needed is an email address.
- To permit granular permissions, for different levels of access. Users can read, write, or administer data shares. See share permissions for more details.
- To provide a variety of file transfer protocols to fit a variety of different scenarios. Bioshare currently supports file transfers via the browser, SFTP, rsync, and wget.
- To provide a user friendly interface, which makes searching, browsing, and viewing/downloading files easy.

CHAPTER 2

User Guide

Below are links to pages which describe the essential functions of Bioshare.

Home



Share list

The home page lists all of the shares that you have access to. This includes both shares that are owned by by you, as well as those that have been shared with you.

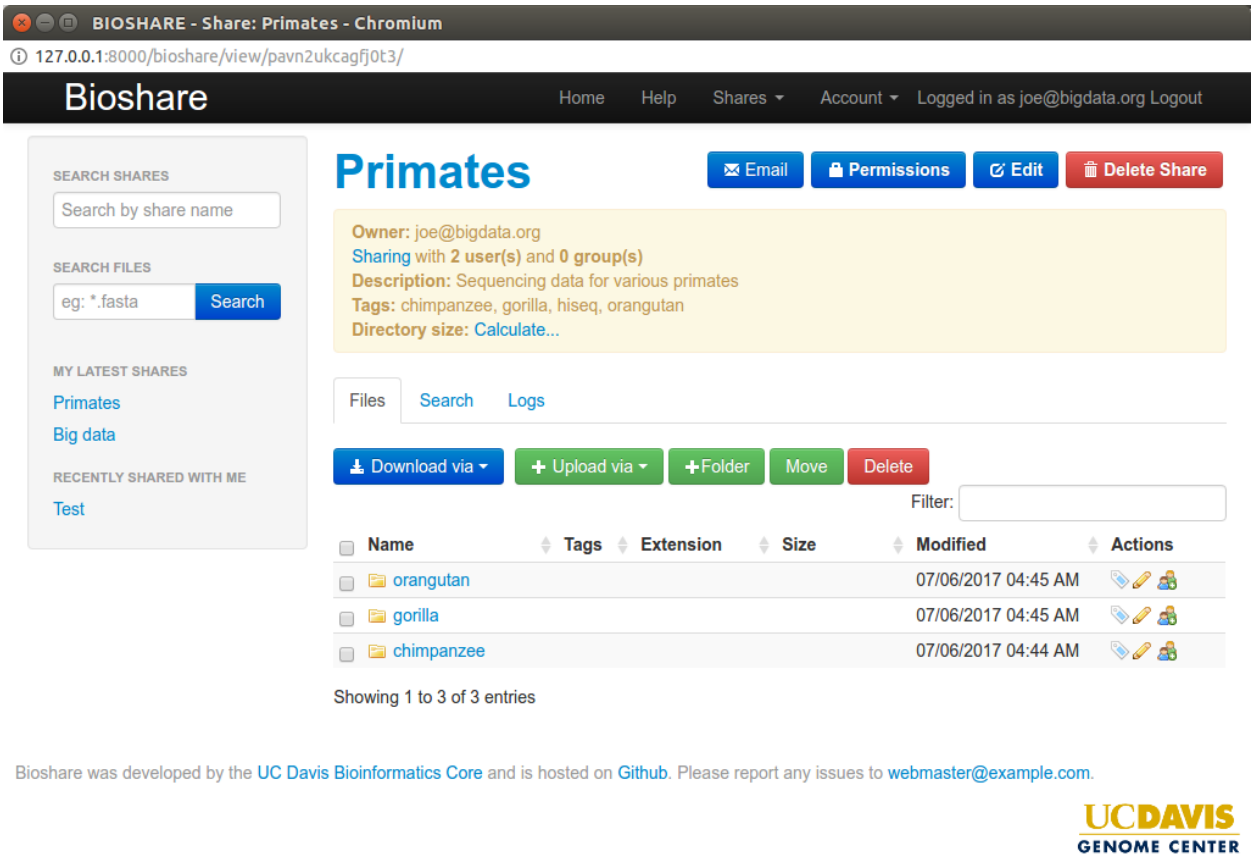
There are several fields that can be displayed when listing available shares, and they can be shown/hidden by checking the box for the corresponding field name under "Fields".

It is possible to sort and/or filter on many of the fields. To sort, simply click on the table header. To sort on multiple fields, hold control as you select the fields that you want to sort on. Clicking a header multiple times will change the direction of the sort. To filter, type some text in the input field under the header. The results will update automatically as you type. If there are too many results, they will be split into multiple pages.

Sidebar

The sidebar is present throughout the application once logged in. It provides shortcuts to shares that you recently created (My latest shares), or shares that have recently been shared with you (Recently shared with me). In addition, you can quickly search for a share by name using the "Search Shares" autocomplete input, or even search for specific files using the "Search Files" search form. Be aware that if you have access to large amounts of data, the file search can be extremely slow and hinder web server performance.

Share



This is the main interface of Bioshare. Here you can view, search, transfer files, and more. A description of the various functionalities follows.

Share information

At the top of the interface are links which indicate the share's name, as well as the full path to the current subdirectory, if applicable. Clicking on the links will bring you to that particular directory in the share. Below the share actions is a box containing general information about the share, including the owner, who it is being shared with, a description, tags, and a link that can be clicked to calculate the size of the current directory (including sub-directories).

Share actions

To the right of the share name, there are a few buttons that perform actions for the entire share:

- Email: Opens a dialog that permits the user to quickly send an email to all accounts that have access to the share.
- Permissions: Redirects to the page for managing share permissions.
- Edit: Redirects to a form that allows the user to update share details.
- Delete Share: Delete the share and all it's contents. Confirmation will be required prior to deletion.

File list

At the bottom of the interface is the list of files and directories for the share. Along with the name, various attributes like size and modification date are available. It is possible to sort files by clicking on the headers, as well as filter them using the input box just above the list. Clicking on the filename will download that file. If it is a directory, the link will take you to a directory listing.

File/Directory action buttons

- Download: Download multiple files or directories, instead of individually clicking on links. Currently supported methods are zipfile (up to a certain size), SFTP, rsync, and wget.
- Upload: Upload multiple files or directories. Currently supported methods are by browser, SFTP, and rsync.
- Create folder: Create a new directory in the current location. Must abide by the naming restrictions for files and directories (alphanumeric characters, periods, underscores, or dashes).
- Move: Move selected files and directories into another location in the same share.
- Delete: Permanently delete selected files and directories.

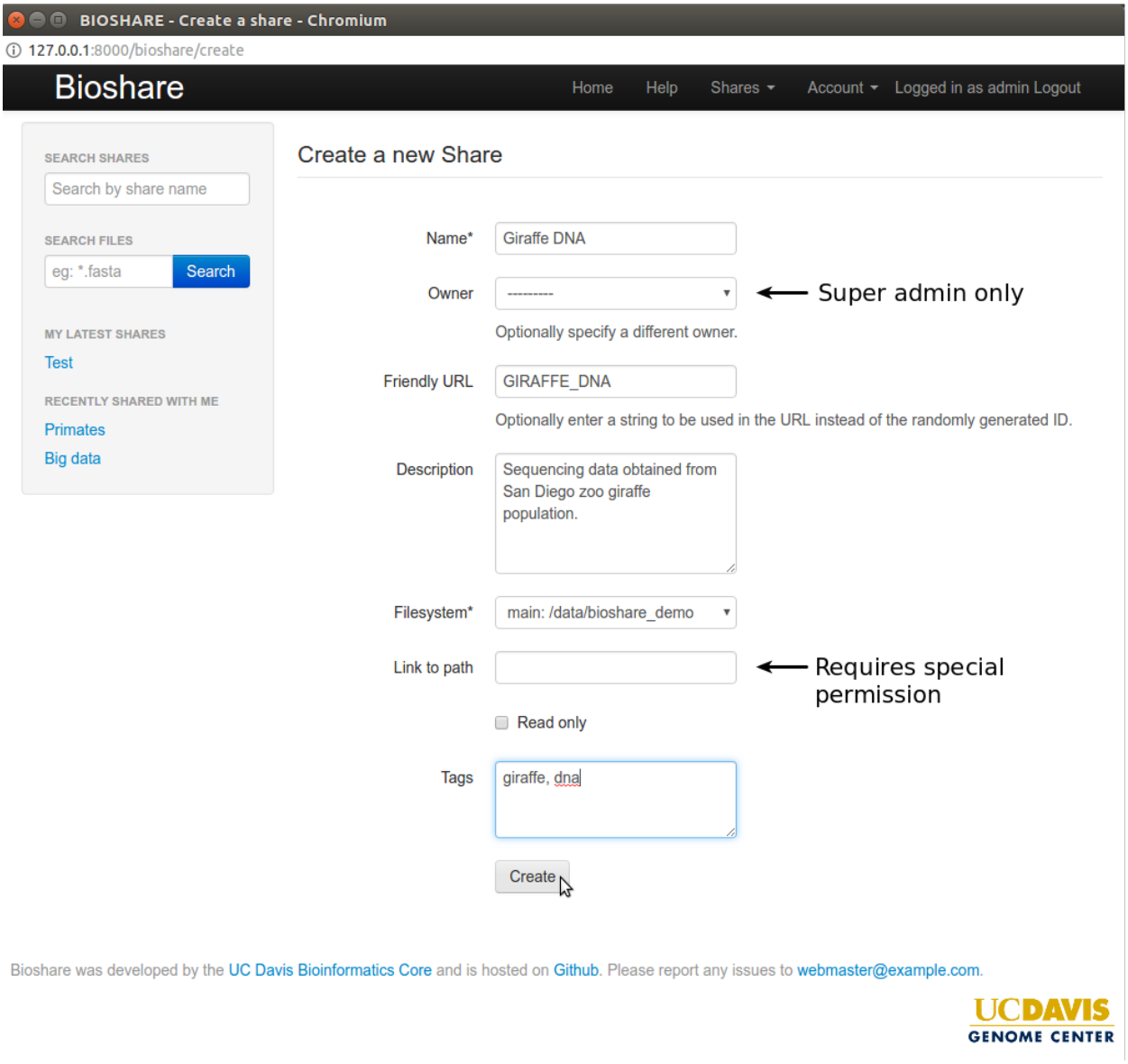
Individual File/Directory actions

Under the "Actions" column, there are a few actions that can be taken on individual files or directories:

- Add/Edit Metadata: Add notes or comma seperated tags
- Modify file/directory name: Just like it sounds. Only use alphanumeric characters, periods, underscores, or dashes.
- Create subshare (share owner only): For the case of a directory, create a share that links to the directory. The subshare will mirror exactly what is in the directory that was reshared. This is useful if you want to assign additional permissions for that directory only.

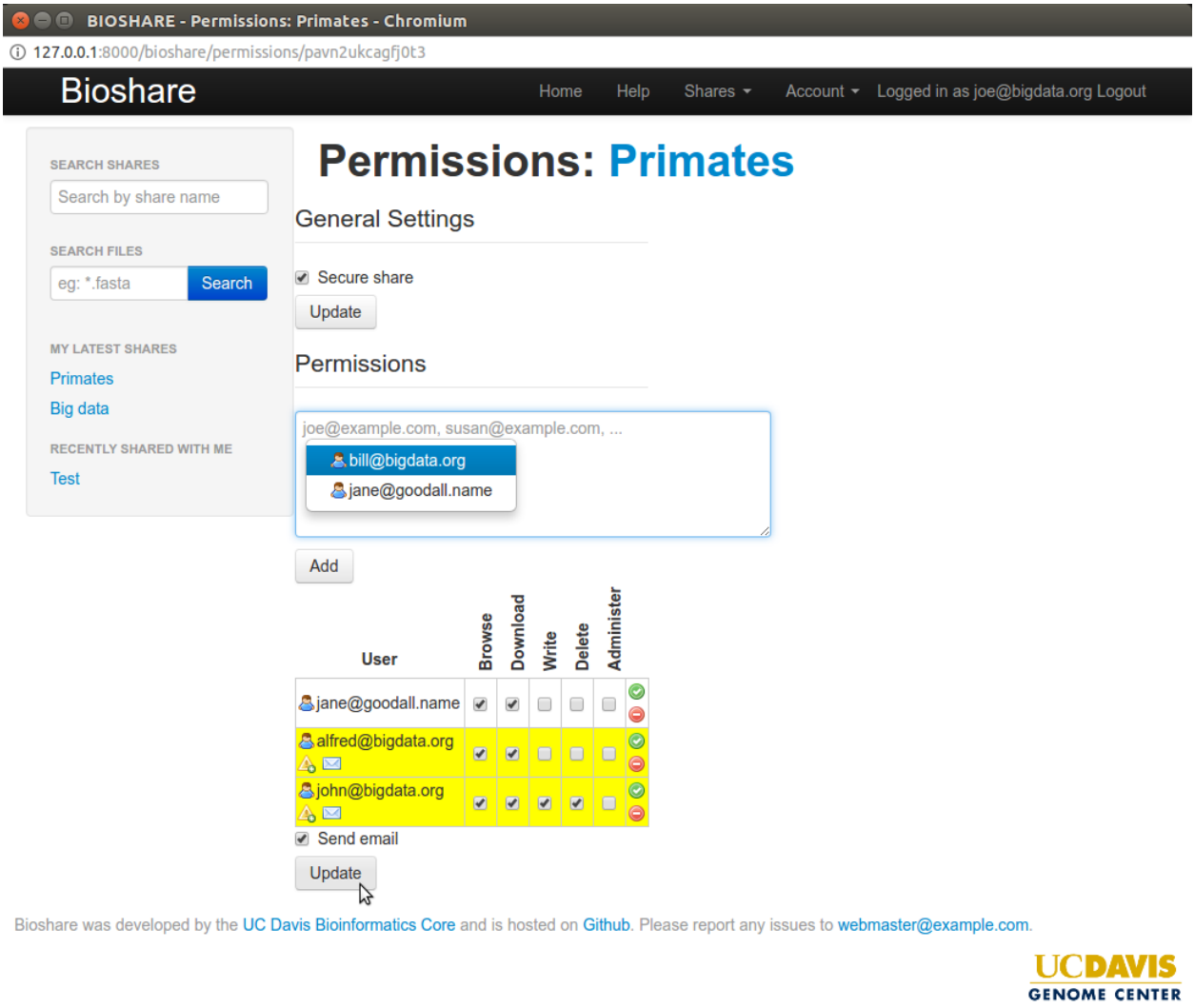
Creating a share

If you have permission to create a share (not all users have permission - many users are just collaborators), it only requires filling out a few mandatory fields. To create a share, from the navigation bar go to "Share -> Create". A description of the process is below:



Name (required): Enter a name for the share. A meaningful name will make it easier to search for.**Owner (super admin only)**: A super admin can assign ownership of the share to another user.**Friendly URL**: Optionally add a text string to use instead of the random ID in places. This will show up in some URLs, as well as the directory listing for SFTP.**Description**: Optional.**Filesystem (required)**: Choose which filesystem to create the share on. Bioshare supports multiple filesystems or mount points.**Link to path (requires special permission)**: It is possible to link to an existing directory on the server. For obvious reasons, this permission is restricted. Additionally, the directoy must fall under a whitelist configured in the config.py file.**Read only**: If selected, writing and deleting from the share will be disabled. Usually this would only be selected when modifying the share after it already contains data.**Tags**: Optionally catagorize the share by listing comma seperated tags.

Share permissions



It is from this page that share admins can manage permissions for the share. Here, they can share with users or groups, and assign them fine tuned permissions.

General Settings - Secure share

Check this box if you want users to have to authenticate to view/download files. If it is not checked, anyone with the URL can access the share. If checked, a user must be listed in the permissions with browse and download permissions in order to view the share.

Regardless of this setting, users have to be authenticated and have appropriate permissions in order to write, delete, or administer the share.

Permissions

Using the textarea, enter email addresses seperated by commas for people to be added to the share. If sharing with a group, you can use the group name instead of an email address. The group name must be prefixed by "Group:",

for example, "Group:my group". If you have previously shared with that user or group, it may try to autocomplete the address or group name. After having entered users and/or groups, click on the "Add" button. They will then be added to the list of users with permissions below. **They do not yet have permission.** In order to finish assigning permissions, it is necessary to click on the "Update" button below the list of user permissions. A description of that process follows.

Permission list

All users and groups with permissions for the share are listed in this table. Each row represents a user or a group, and the columns are as follows:

- User (or group): The email address or group name. An icon will be shown here to warn if the user is being added. An icon will also be shown indicating whether they will receive an email.
- Browse: This base permission is required to be able to even view the share.
- Download: Required to be able to downlod files. If browse and download are selected, the user will also be able to download via SFTP or rsync.
- Write: Required to create/upload/rename files and directories.
- Delete: Required to delete files and directories. Additionally, this permission is required in order to upload via rsync.
- Administer: Allows the user or group to perform administrative tasks, including modifying the share's permissions.
- Select/Deselect all: Convenience buttons for rapidly selecting or deselecting permissions.

Below the list, there is a "Send email" checkbox. If selected, any user/group who is newly added to the share permissions will be sent an email with a link to the share.

Users who do not yet exist in the system will be sent an email with their new user credentials and a link to the share, regardless of if the box is checked.

If a user or group has had their permissions changed, that row will be highlighted in yellow. Please note, that no permissions will be updated until the "Update" button is clicked. Once the update button is clicked, a confirmation message should appear in the upper right corner of the screen, and no more rows should be highlighted in yellow.

Downloading/Uploading

There are multiple ways to download or upload files within Bioshare. The simplest is generally by using the browser. Here, files can simply be downloaded by clicking on a link, or by selecting a list of files and directories and downloading them as a zip file. Uploads are simple as well, using the same mechanisms that most web applications use for selecting files to upload.

There are, however, times where transfering files using the browser is undesirable. For large or numerous files, browsers do not handle failure well. It can be difficult or time consuming to know where to things back up. In these cases, tools like rsync, SFTP, or wget prove to be useful.

Another scenario where using the browser is impractical is when trying to transfer to or from another server. If for example, I am browsing files on Bioshare using my laptop, but I ultimately want them available on my computing cluster, I would first need to download the files to my laptop, and then upload them from my laptop to the cluster. This is both inefficient and time consuming. A much more efficient method is to log in to the server where I want to store the data, then using command line tools (rsync, SFTP, wget), transfer the data directly.

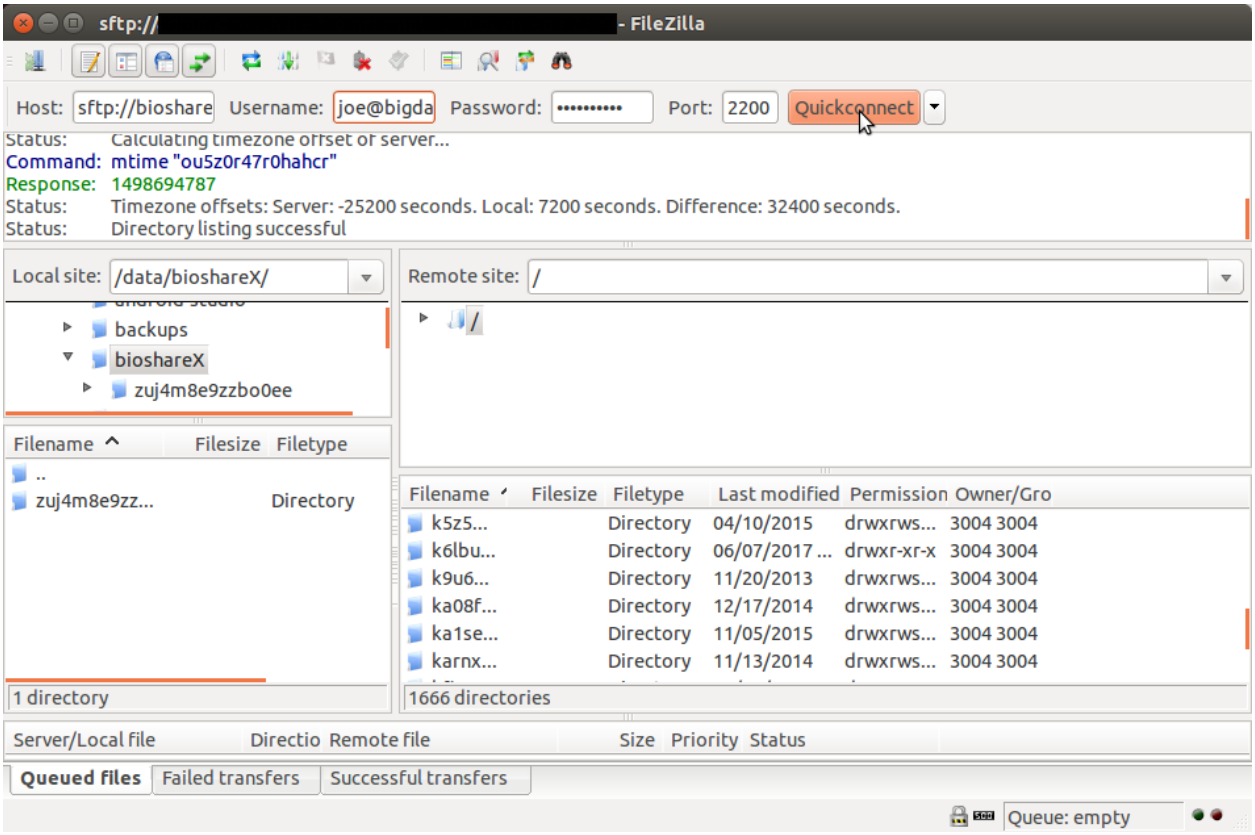
SFTP

SFTP is something of a cross platform standard for secure file transfer. It is available on most any operating system either as a GUI, or as a command line application. Bioshare supports both downloads and uploads using SFTP.

Using FileZilla

As far as graphical applications that support SFTP go, I often suggest FileZilla, as it is free, open source, and available on Windows, Mac, and Linux. There are plenty of other options as well, such as Cyberduck, but not all support all platforms.

The simplest way to connect to Bioshare using FileZilla is simply by using the "Quickconnect" parameters at the top of the window. Here, you need to enter the host (ie: sftp://bioshare.bigdata.org), username (same as you log into Bioshare with), password (again, same as you log in with), and port (depends on how Bioshare was configured). Once entered, click "Quickconnect" and you should see some details below as FileZilla attempt to log in. After successful authentication, you'll have a list of all of your available shares on the right side of the window. You can simply drag and drop files or directories between Bioshare and your local machine.



Using the command line (Linux/Mac)

Again, this is a useful option when trying to transfer files to/from a remote server. Connecting using sftp is quite simple, and will look something like (depending on port number, username, etc):

```
sftp -P 2200 'joe@bigdata.org'@bioshare.bigdata.org
```

To view a list of shares, use the "ls" command. If the share has a "friendly URL", it will be listed by that instead of the random directory. To change into a directory, use "cd DIRECTORY_NAME". To download files, use "get FILENAME". If it is a directory, use "get -r DIRECTORY_NAME". For uploading files, use "put /local/path/to/file" or for a directory "put -r /local/path/to/directory". These are a few very simple examples. For more details on how to use sftp on the linux or mac command line, see the manual.

Rsync (Linux/Mac)

"Rsync is an open source utility that provides fast incremental file transfer." In other words, this is probably the most efficient option for transfering data. Additionally, rsync is able to identify which files have already been transfered, and if the source has been updated since. In this way, rsync will only transfer the files that have been changed, which is important if regularly updating large datasets.

Rsync is, however, the least simple transfer method to setup. Unlike SFTP, Bioshare cannot use your username and password to authenticate you. It instead relies on the use of SSH keypairs. And while it is possible to configure a Windows system to use rsync, it is not recommended.

Bioshare uses SSH keypairs in much the same way as Github. Github has good documentation on how to use SSH keypairs for authenticating to Github, and is worth a read.

Setup

Here are the basic steps to setting up SSH keys with Bioshare.

1. Make sure you have an SSH keypair:

```
ls -al ~/.ssh
```

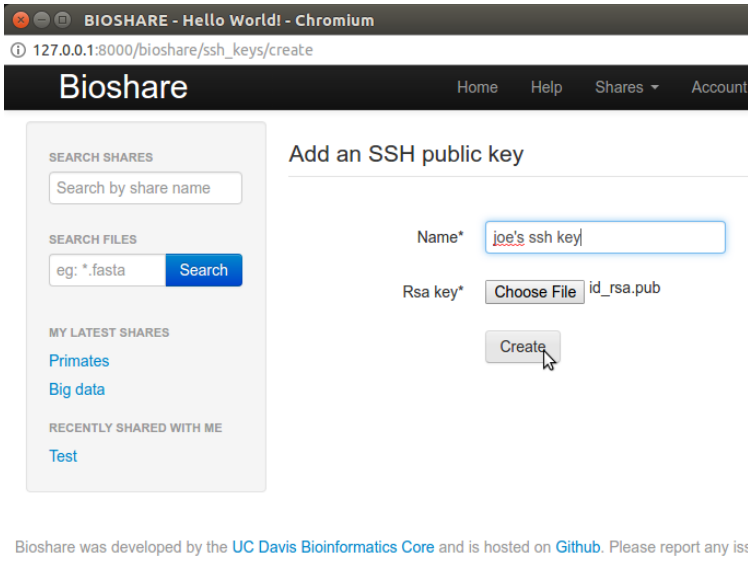
Look for a pair of files called "id_rsa" and "id_rsa.pub". Those would be your private and public keys, respectively. If you have them, you can skip to step 3.

1. Create an SSH keypair if you didn't find one you want to use in step 1:

```
cd ~/.ssh
ssh-keygen -t rsa -b 4096 -C "your_email@example.com"
```

When asked for a passphrase, make sure to enter one. You'll be asked to enter this each time you use the keypair, which will keep it secure.

1. Upload your public key, i.e. "id_rsa.pub" to Bioshare.
   (a) Once logged into Bioshare, under the menu go to Account->SSH Keys. If you have any public keys uploaded, they'll be listed here, along with the possibility to delete them.
   (b) Click on the button to upload a public key.
   (c)
   - Select a name for the public key. This serves only to remind you of what key you uploaded.
   - Select the public key (id_rsa.pub) that you will be using. It should be in your home directory, in the .ssh directory. If you can't find the .ssh directory it may be that your browser isn't listing files/directories that start with a ".". You can always copy the public key somewhere more convenient for upload in this case.



- Click create. You should now see your key listed.

Rsyncing

Bioshare will generate the appropriate rsync command for you from any given share or share directory. A typical command to download files from Bioshare would be:

```
rsync -vrt bioshare@bioshare.bigdata.org:/RANDOM_SHARE_ID/ /to/my/local/directory
```

The general format for uploading files is:

```
rsync -vrt --no-p --no-g --chmod=ugo=rwX /path/to/my/files bioshare@bioshare.bigdata.
↪org:/RANDOM_SHARE_ID/
```

Troubleshooting rsync

It is asking for my password, and failing to authenticate

If your SSH key is set up properly, Bioshare should not ask for a password. Here are a couple things you should check for, assuming you already set up your SSH keypair as described in the setup section above:

1. Is the private key corresponding to public key you uploaded to Bioshare on the current system under ~/.ssh/id_rsa? Some people may try to SSH into a server from their desktop, then rsync from there. In this case, the server you are rsyncing from may not have your SSH private key available. To resolve this, you can either:

   - Copy your private key to the server in the appropriate place (~/.ssh/). Then, again assuming your key is named id_rsa, add it to your SSH agent:

   ```
   ssh-add ~/.ssh/id_rsa
   ```

   If for some reason you get an error message like, "Could not open a connection to your authentication agent.", you'll need to start the SSH agent first:

```
eval "$(ssh-agent -s)"
ssh-add ~/.ssh/id_rsa
```

2. Your SSH private key is on the machine you are rsyncing from, but the ssh agent isn't started. To check if the agent is running, enter:

```
ssh-add -L
```

If you got an error like "Could not open a connection to your authentication agent.", you need to start it and add your key:

```
eval "$(ssh-agent -s)"
ssh-add ~/.ssh/id_rsa
```

1. Okay, your key is on the server, and your SSH agent is running. Are you using the right key? Run:

```
ssh-add -L
```

Make sure that in the output a key matches one of the ones on the "Bioshare SSH Keys" (see Setup section above) page. If not, use ssh-add to add the corresponding private key:

```
ssh-add ~/.ssh/id_rsa
```

When downloading, I'm getting an error similar to "handle_rsync exception: User email@example.com cannot read from share"

You do not have permission to download from the share. Ask the owner of the share or a share administrator to give you download permissions.

When uploading, I'm getting an error similar to "handle_rsync exception: User email@example.com cannot write to share"

You do not have permission to upload with rsync to the share. Ask the owner of the share or a share administrator to give you both write and delete permissions. Both of these are necessary to upload with rsync.

CHAPTER 3

Indices and tables

- genindex
- modindex
- search